MOTION IMAGERY STANDARDS BOARD	MISB TRM 0909.3
Technical Reference Material	**19 April 2011**
Constructing a MISP Compliant File/Stream	

1 Scope

This document provides an example MISP (Motion Imagery Standards Profile) compliant Full Motion Video (FMV) file/stream constructed from technologies approved by the Motion Imagery Standards Board. The intent is to identify the key elements that comprise a compliant FMV file/stream; not to show all possible combinations of available technologies. Finally, this document is a good starting point for those new to the practices of the MISB for it indicates further documentation useful in adhering to MISB practices.

2 References

[1]	MISB RP 9720b	*MISM, High Definition Motion Imagery*, MISP
[2]	MISB EG 0904	*H.264 Bandwidth/Latency/Quality Tradeoffs,* Sep 2009
[3]	MISB STD 9710	*High Definition Television Systems (HDTV)*, MISP
[4]	MISB RP 0605	*Inserting Time Code and Metadata in High Definition Uncompressed Video*, May 2008
[5]	MISB STD 9715	*Time Reference Synchronization,* MISP
[6]	MISB RP 0603	*Common Time Reference for Digital Motion Imagery Using Coordinated Universal Time (UTC),* Aug 2006
[7]	MISB STD 9723	*Compressed High Definition Advanced Television (ATV) and Associated Motion Imagery Systems*, MISP
[8]	MISB EG 0802	*H.264/AVC Coding and Multiplexing*, May 2009
[9]	MISB STD 0107	*Bit and Byte Order for Metadata in Motion Imagery Files and Streams*, Oct 2001
[10]	MISB STD 9713	*Data Encoding Using Key-Length-Value*, MISP
[11]	MISB STD 0807	*MISB KLV Metadata Dictionary*, Jun 2010
[12]	MISB EG 0607	*MISB Metadata Registry and Processes*, Dec 2008
[13]	MISB TRM 1006	*Key-Length-Value* (*KLV) Users Guide*, May 2010
[14]	MISB STD 0902	*Motion Imagery Sensor Minimum Metadata Set*, Jun 2010
[15]	MISB STD 0601	*UAS Datalink Local Metadata Set*, Mar 2010
[16]	MISB STD 0102	*Security Metadata Universal and Local Sets for Digital Motion Imagery*, Sep 2009
[17]	MISB STD 9701	*MPEG-2 Transport Stream*, MISP
[18]	MISB STD 9708	*Imbedded Time Reference for Motion Imagery Systems*, MISP
[19]	MISB STD 0604	*Time Stamping Compressed Motion Imagery*, Mar 2010
[20]	IETF Draft	*RTP Payload for SMPTE 336M Encoded Data*, draft-ietf-avt-rtp-klv-00.txt, Apr 2010
[21]	MISB RP 0804	Real Time Protocol for Full Motion Video, May 2010

[22] MISB STD 9703 *Digital Motion Imagery, Uncompressed Baseband Signal Transport and Processing*, MISP

[23] MISB RP 9717 *Packing KLV Packets into MPEG-2 Systems Streams*, MISP

[24] MISB RP 0101 *Use of MPEG-2 Systems Streams in Digital Motion Imagery Systems, Jan 2011*

[25] MISB TRM 1007 *Surfing the MISP*, May 2010

[26] ISO/IEC 13818-1 *Information Technology – Generic coding of moving pictures and associated audio information, Part 1: Systems*, 2007

[27] SMPTE RP217-2001 *Non-Synchronized Mapping of KLV Packets into MPEG-2 System Streams*

[28] SMPTE EG 40-2002 *Conversion of Time Values Between SMPTE 12M Time Code, MPEG-2 PCR Time Base and Absolute Time*

3 Acronyms

AAF	Advanced Authoring Format
AVC	Advanced Video Codec (H264)
DAM	Digital Asset Management System
DTS	Decoding Time Stamp
EG	Engineering Guideline
ES	Elementary Stream
FMV	Full Motion Video
Gb	Giga-bits (10^9)
IRSM	Infrared System Matrix
Kb	kilo-bits
kB	kilo-bytes
KLV	Key-Length-Value
LVSD	Large Volume Streaming Data
Mb	Mega-bits (10^6)
MI	Motion Imagery
MISM	Motion Imagery System Matrix
MISP	Motion Imagery Standards Profile
MXF	Material Exchange Format
PED	Processing-Exploitation-Dissemination
PES	Packetized Elementary Stream
PID	Packet IDentifier
PTS	Presentation Time Stamp
RP	Recommended Practice
RTP	Real Time Protocol
RTSP	Real Time Streaming Protocol
SDI	Serial Data Interface
SMPTE	Society for Motion Picture and Television Engineers
STD	Standard
TRM	Technical Reference Material
TS	MPEG-2 Transport Stream
UTC	Coordinated Universal Time
VANC	Vertical Ancillary Data
VBI	Vertical Blanking Interval

4 Introduction

The definition of MISP Compliance provided in MISP 6.0 is as follows:

MISP COMPLIANCE Definition:

"Motion Imagery Standards Profile (MISP) compliance is based upon compliance to a specified approved version of the MISP (e.g. MISP Version (V) 4.4, MISP V4.5, etc.). The motion imagery system supplier specifies the MISP version for which it is seeking compliance along with four qualifications:

1. The MISM-Level(s) that the video compression is to operate within (across)
2. The metadata STD/RP/EG implemented
3. The transport/file format STD/RP/EG for transport/storage of the motion imagery
4. The timing/synchronization STD/RP/EG implemented

MISM levels are as defined per the MISP version specified by the system supplier. All signals tested are assumed digital. Supported video compression includes MPEG-2, MPEG-4 Part 10 (i.e. AVC or H.264) and motion JPEG 2000. Metadata is tested for compliance to the specified version of the MISP and respective STDs/RPs/EGs. MISP compliant systems shall produce metadata elements from Standard 0601 or EG 0104 (legacy systems only), optionally using metadata keys from MISB Standard 0807, SMPTE RP 210, and other MISB RPs/EGs of their choice. The minimum metadata set for Standard 0601 is given by Standard 0902. In addition, Security metadata shall comply with MISB Standard 0102. Supported file formats include MPEG-2 Transport Stream (TS), Real Time Protocol (RTP), Material Exchange Format (MXF) and Advanced Authoring Format (AAF). Furthermore, if the motion imagery system uses MXF/AAF it shall comply with Standard 0301. Draft RPs/EGs will not be tested until approved by the MISB. Note: While it is strongly advised to reference the latest version of the MISP in assuring full compliance, adherence to MISP Version 4.4 and subsequent versions fulfill the requirements for compliance testability."

A MISP-compliant file/stream for FMV (Full Motion Video) must have the following three components with timing/synchronization implicit in the metadata and/or media container:

1. **Motion imagery**: uncompressed or compressed (the essence)
2. **Metadata**: Key-Length-Value form (e.g. mission, security, optional PED data)
3. **Media Container**: approved package that carries MI, metadata, or both. In a MPEG-2 transport stream (TS) both MI and metadata must be present. In Real-Time Transport Protocol (RTP), MI and metadata are independently carried as separate streams with appropriate timing that associate the individual media streams.

Files and streams consist of compressed imagery and metadata appropriately supplied with timing. A "file" is generally considered a complete media package, such as a video clip stored on a server. A "stream" could be a live event transmitted directly from a sensor, or a pre-recorded file that is subsequently played out in continuous fashion. Both files and streams are composed of motion imagery and metadata that may be either multiplexed together into one unified transport stream container or each packaged within separate RTP containers.

5 File/Stream Construction - example

As way of illustration, the various fundamental building blocks that comprise a compliant file/stream are shown in Figure 1. Note that this may not indicate how these functions are implemented or what in what order some operations may occur.

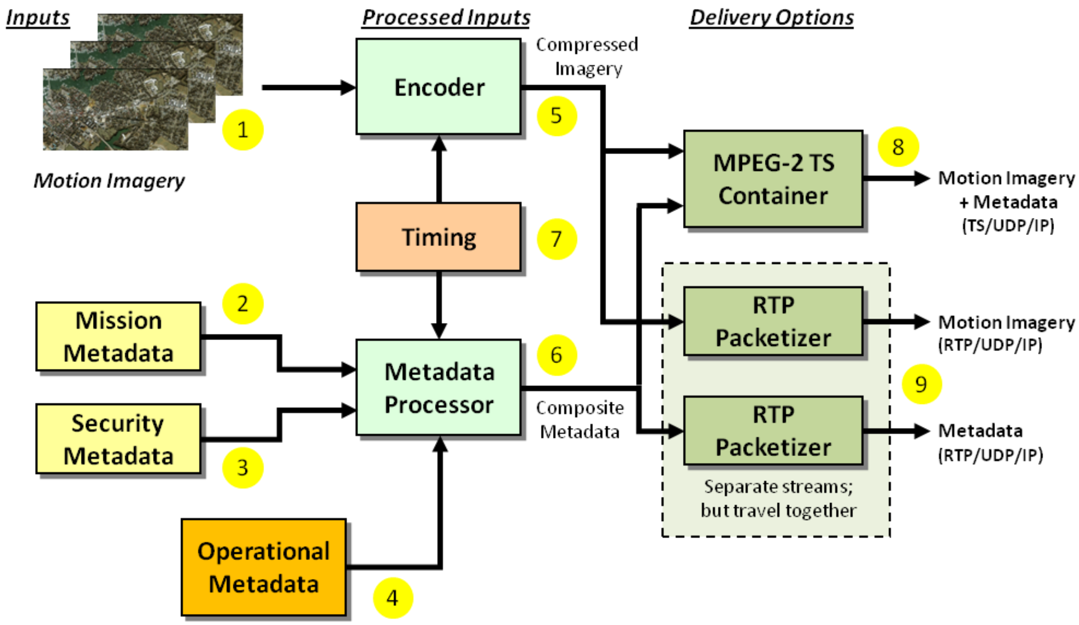


Figure 1: A General Structure for Constructing a MISP-Compliant File/Stream

1. A motion imagery sequence (e.g. FMV) is produced by a sensor
2. Mission Metadata (geo position, sensor angles, etc.) is output by the mission computer
3. Security Metadata (required) is supplied
4. Optional metadata (e.g. chat, audio, annotation) may be added
5. An encoder (if desired) compresses the motion imagery sequence (e.g. H.264, MPEG-2)
6. All metadata is multiplexed together in preparation for the MPEG-2 TS container
7. Absolute Time Stamps are applied to the MI and Metadata
8. MI and Metadata are combined into a unified MPEG-2 Transport Stream Container
9. MI and Metadata may be packaged into separate RTP streams

Table 1 identifies the functional elements needed to create a MISP compliant motion imagery file. The dotted lines indicate that these data may be carried either independently or together. The MISB is very specific in defining methods for data transport, allowed compression types, metadata syntax, semantics and elements, timing standards and relationships, and data packaging using standardized containers. These details are more thoroughly defined within the MISP and associated documents.

Input Data Package	Media Components	Relationship	Media Container
Transport Protocol	Compressed Motion Imagery	Timing/Multiplexing	Transport Protocol
Transport Protocol	Metadata		Transport Protocol

MISP Compliant File / RTP Streams

Table 1: Media Elements, Timing, and Container

Table 2 shows that motion imagery may be supplied over a SDI (serial digital interface) - *the transport protocol* to an H.264/AVC encoder-*media component*. The metadata-*media component* may be supplied within the same SDI signal, or via some other input protocol means, such as Ethernet, or RS-232. The relationship between the two media component types is maintained through timing assigned and contained within each media component. The two media components may then be packaged together, or as individual single components, depending on the desired protocol.

Input Data Package	Media Components		Relationship	Output Data Package
Transport Protocol	**Compressed MI**		**Timing/Multiplex**	**Transport Protocol**
HD-SDI SDI Other	H.264/AVC MPEG-2		MPEG-2 TS RTP	***File*** MPEG-2 TS AAF/MXF
	Metadata			***Stream*** MPEG-2 TS RTP
	Mission	***Security***		
	KLV			

Table 2: Specific Technologies for Electro-optical and Infrared Imagery

6 Example Implementation: *A Digital HD FMV Sensor System*

Figure 2 shows an example high-definition digital sensor collection system with the processing steps described in Figure 1. In Appendix A, an example legacy analog sensor system is shown. Various MISB standards and practices that are relevant to each step are indicated for reference.

Step 1: *FMV Source – sensor and channel constraints*

Since the FMV source is high definition (HD) video it is first prudent to review the appropriate Motion Imagery System Matrix (MISM) table in the MISP to determine if the overall system can support the bandwidth to meet the CONOPS. **MISB RP 9720b - MISM, High Definition Motion Imagery [1]** indicates that there are three levels that support high definition.

RP 9720b spans image spatial resolutions from 1920 by 1080p at 60 frames-per-second to 1280 by 720p at 30 frames-per-second. The compression type, MPEG-2 or H.264/AVC, dictates the data rate to support a desired level of quality based on the application. In this case, the data rate ranges from a maximum of 2.4 Gb/sec in Level L11 (for uncompressed HD) to a nominal bandwidth of 6 Mb/sec (H.264/AVC) in Level 9H. **MISB EG 0904 - H.264 Bandwidth/Latency/Quality Tradeoffs [2]** offers suggested methods to meet system bandwidth constraints by reducing the spatial image size (scaling), reducing the image field of view (cropping), or reducing the temporal rate.

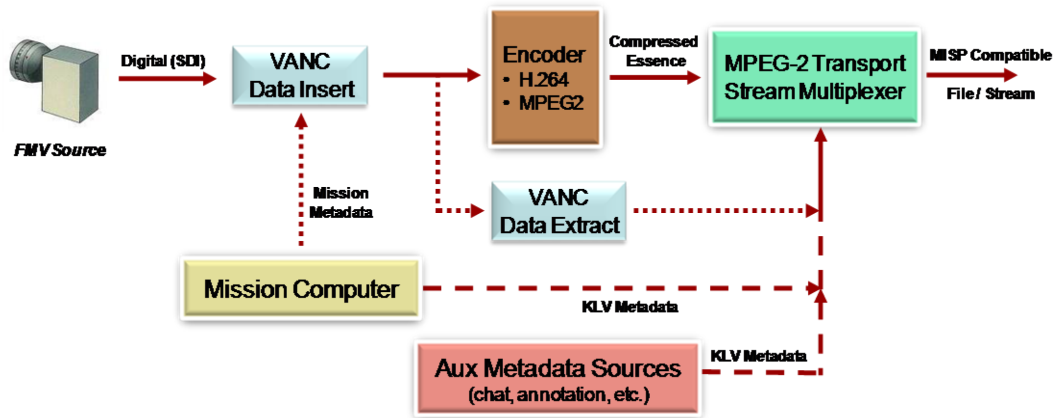


Figure 2: Digital FMV Sensor System

Step 2: *VANC Data Insert – sensor interface*

The next step is to understand the nature of the sensor's interface. If the interface adheres to an industry standard then the data format and its usage is well defined. In this example, the SMPTE HD-SDI (High Definition Serial Digital Interface) is an industry standard (see **MISB STD 9710 - High Definition Television Systems (HDTV) [3]** for more).

The digital FMV data is transported from the sensor to the encoder by HD-SDI. In the HD-SDI signal, a reserved data space called the VANC (Vertical Ancillary Data Space) is available within every video frame to insert non-video data, much like the Vertical Blanking Interval (VBI) for NTSC. In this example, mission metadata is inserted into the VANC for carriage to as the encoder; the encoder will subsequently extract this metadata from the VANC.

Also, placed within the VANC is a time stamp. This absolute time stamp is based on UTC (Coordinated Universal Time). This time stamp will be used to populate the time stamp metadata element in the metadata stream, and can be used to populate the required time stamp in the compressed video essence. For more on the VANC and its capabilities see **MISB RP 0605 - Inserting Time Code and Metadata in High Definition Uncompressed Video [4]**. For time stamping information see **MISB STD 9715 - Time Reference Synchronization [5]** and **MISB STD 0603 - Common Time Reference for Digital Motion Imagery Using Coordinated Universal Time (UTC) [6]**.

Step 3: *Encoder – motion imagery compression*

The high definition motion imagery is compressed using one of the allowed compression types for video, which include H.264/AVC and MPEG-2 (see **MISB STD 9723 - Compressed High Definition Advanced Television (ATV) and Associated Motion Imagery Systems [7]**.) The MISB leverages industry standards and practices where possible; however, oftentimes where a standard is very broad, the MISB will seek to limit the standard in order to achieve community interoperability. In particular, **MISB EG 0802 - H.264/AVC Coding and Multiplexing [8]** provides parameters and application-specific modes within the H.264/AVC standard that are deemed more optimal within our community of practice.

Step 4: *Metadata – mission, security, and auxiliary*

Metadata is collected by the mission computer, or optionally input from other sources, such as by an operator. If metadata is inserted into the SDI VANC then that data must be extracted and input to the transport stream multiplexer. Otherwise, the metadata may come from the mission computer, and be input to the multiplexer directly.

Regardless of the metadata source all metadata must abide by the MISB standards for syntax and semantics. SMPTE KLV (Key-Length-Value) is how all MISB metadata is encoded. This binary format was chosen because of its efficiency, extendibility and decodable robustness. Certain MISB standards govern the format of the metadata (see **MISB STD 0107 - Bit and Byte Order for Metadata in Motion Imagery Files and Streams [9]** and **MISB STD 9713 - Data Encoding Using Key-Length-Value [10]**). The MISB maintains a listing of all metadata keys in **MISB STD 0807 - KLV Metadata Dictionary[11]** with a corresponding document **MISB EG 0607 - MISB Metadata Registry and Processes [12]** that describes the fundamentals of the metadata dictionary. A tutorial on KLV structures and rules is found in **MISB TRM 1006 - Key-Length-Value (KLV) Users Guide [13]**.

While many KLV elements supported by the MISB describe important mission data, not all are mandated to meet MISP compliance. Certain mission metadata and security metadata must, however, be present.

Mission Metadata

MISB STD 0902 - Motion Imagery Sensor Minimum Metadata Set [14] specifies the mandated set of KLV elements that characterize many of the dynamic parameters collected during a mission. This set is drawn from a more complete set defined in **MISB STD 0601 - UAS Datalink Local Metadata Set [15]**. The 0902 minimum set enables basic discovery and retrieval functionality in exploitation. While certain mission metadata may be updated at the motion imagery frame rate, "Metadata elements contained in the Minimum Metadata Set **shall** be reported no less than once every thirty seconds (30) under all circumstances" (see STD 0902).

Security Metadata

The security metadata set is defined in **MISB Standard 0102 - Security Metadata Universal and Local Sets for Digital Motion Imagery [16]** and must be present within the metadata file or stream to be MISP compliant. It is recommended that this KLV local data set be placed early in the file/stream to help determine the security level of the data early without searching through the entire dataset, and that its repetition within the file/stream occur at least every 10 seconds (STD 0902). *A point of clarity: the required elements of the local security set as defined in Standard 0102 are identified within Standard 0902. Thus, choosing to define security using Standard 0902 satisfies the requirement for MISP compliance.*

Guidelines for how to link security metadata to MPEG elementary streams, programs within an MPEG-2 transport stream, an entire transport stream, more than one transport stream, and linking to other metadata elements can be found in **MISB Standard 0102 §5.4**.

Auxiliary Metadata (optional)

Other types of metadata may be inserted at the sensor, or anywhere within the processing/exploitation workflow. Numerous MISB practices for data such as chat, annotation, and audio are documented in various MISB RP's and EG's. All such metadata must conform to the KLV format. Care must be taken to allow for the additional overhead in bandwidth that these types of metadata may incur. Refer to the MISB website for more information on these various metadata constructs.

Frame Co-incident versus Non-Frame Co-incident Metadata

Metadata that may be highly correlated or co-incident to a motion imagery frame must contain an absolute time stamp, as well as, the motion imagery it pertains to. Each of these time stamps in turn reference UTC, and thus through that reference can reference one another. It is important to distinguish that such absolute timing between essence types is different than the way they may be transported within an MPEG-2 Transport Stream, for example.

Step 5: *MPEG-2 Transport Stream Multiplexer – timing and transport*

With the high definition video in compressed form (compressed essence, Figure 1) and the KLV metadata populated the job is now to package the two into a container or wrapper for delivery (or storage). In the example of Figure 1, the two media types are combined into a MPEG-2 transport stream; see **MISB STD 9701 - MPEG-2 Transport Stream [17]**.

MISB STD 9708 - Imbedded Time Reference for Motion Imagery Systems [18] identifies the three documents RP 0603, STD 0604, and RP 0605 as the DoD/IC/NSG standards for time annotation and imbedded time references for motion imagery systems. Understanding the timing mechanisms allowed under the MISB is critical to successful application and compliance.

MISB STD 0604 - Time Stamping and Multiplexing of Compressed Motion Imagery and **Metadata [19]** specifies where to insert a time stamp into a motion imagery stream, and two for multiplexing or combining metadata with the motion imagery. *This is a document that should be read.*

Time Stamps

Mission time stamps are absolute and represent actual event times. They are based on UTC (Universal Coordinated Time) along with the ASCII identifier MISPmicrosectime (see STD 0604). Mission time stamps in the compressed video elementary stream will be identical to those carried in the metadata if both the video and metadata are captured at the same time. Accurate time stamps aid post-processing and exploitation of an event.

Time stamps are <u>required</u> elements in both the motion imagery and the metadata.

Synchronous/Non-synchronous Metadata Carriage

Two methods for metadata carriage are described in STD 0604; a synchronous method specified under ISO 13818-1 [26] and a non-synchronous method specified under SMPTE RP 217 [27]. Among other differences the ISO method allows for a Presentation Time Stamp (PTS) to be applied to the metadata. This affords better assurance that events in the video and co-incident events in the metadata get presented together at a display, for example. Although the PTS of the video and that of the metadata can be made the same, it is also possible to force the PTS values to align with the absolute time stamps, if desired (see SMPTE EG 40 [28]).

It is important to note that the PTS values are not intended to provide a timing relationship between the motion imagery and the metadata for important exploitation and analysis functions. A PTS is a relative timing mechanism intended to align various media streams together for presentation purposes only.

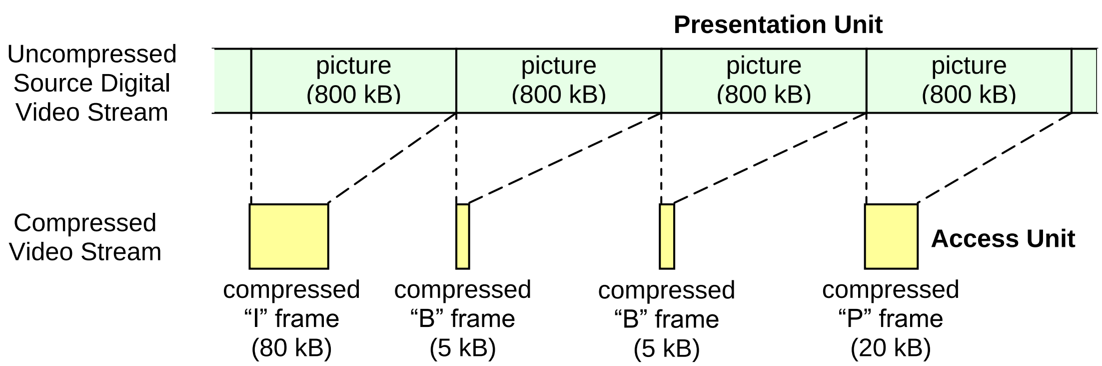
7 The MPEG-2 Transport Stream – an overview

Step A – *the elementary stream*

To understand how data is carried in a transport stream it is instructive to start at the beginning of the process with uncompressed source imagery.

Figure 3 shows a series of uncompressed digital video frames that get compressed.

An MPEG video stream is composed of a series of full, predicted, and bi-directionally predicted frames referred to as I, P, and B frames. I (Intraframe) frames (or full pictures) are coded based on information within a frame independent of surrounding frames and, therefore, generally compress the least (80kB). P (Predictive) frames are predicted from I frames (or other P frames), and thus, compress more than I frames (20kB). Finally, B (Bi-directional) frames are predicted from both I and P frames in a bi-directional fashion using image information from frames both before and after. B frames generally compress the most (5kB); however, their dependence on neighboring frames makes them very susceptible to the propagation of errors in reconstruction. I, P, and B compressed frames are called Access Units. Step A shows a coding structure of one I, two B, and one P frame. This pattern, while not unique, will repeat over the entire coded sequence. Note also that the compressed data values are for illustration only and will differ substantially as a function of image scene complexity and encoder algorithm.



Compression of Presentation Units (frames) to Access Units yields the Video Elementary Stream

Figure 3: Step A: the compression process

Step B – *creating a packetized elementary stream*

Once the source uncompressed imagery is compressed into Access Units it must be prepared for packaging into the MPEG-2 transport stream. Step B in the process (Figure 4) is to create packetized elementary stream (PES) packets of each compressed stream (remember there may be multiple streams of video, audio, metadata, etc) with accompanying information that indicates what is contained in the packet and how it should be decoded. A PES packet header carries this control information.

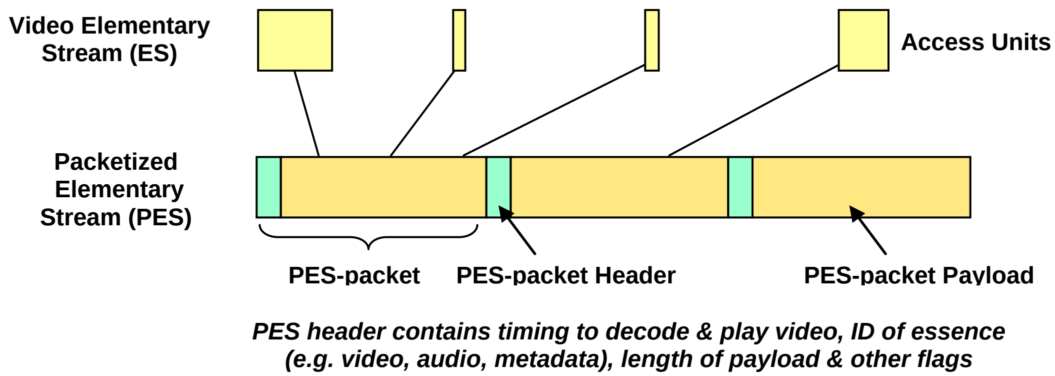


Figure 4: Step B: packetizing the elementary stream

PES packets can vary in length and contain a number of Access Units. Within the PES header a decode time stamp (DTS) aids in decoding the data. A presentation time stamp (PTS) aids in the presentation of the data to the display. The PTS is especially important in synchronizing the different media types (video, audio, metadata) at the display. Appendix B provides a list of various flags and fields in the PES packet.

Step C – *mapping into transport stream packets*

PES packets are mapped into the transport stream itself. Each TS packet is fixed in size (188 bytes). A four-byte TS header contains various indicators, such as a start code for the TS packet, an error indicator for bad packets, a packet ID or PID that identifies what media the packet of data represents. This information is required so that a decoder can decipher the data appropriately. Appendix B lists various flags and fields for the TS header.

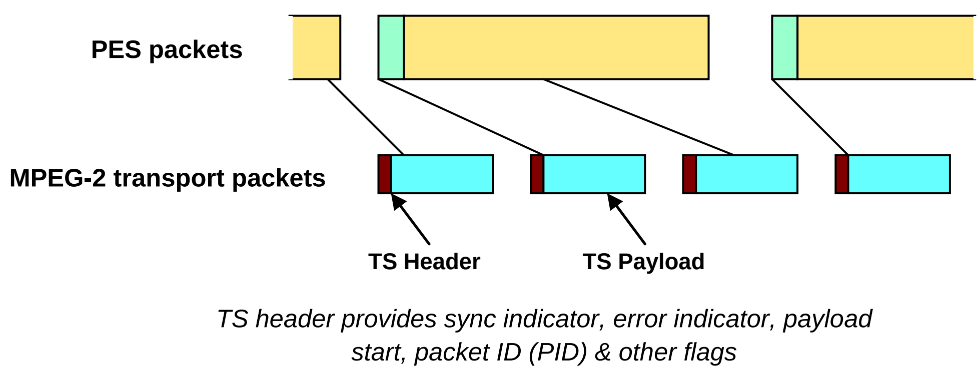
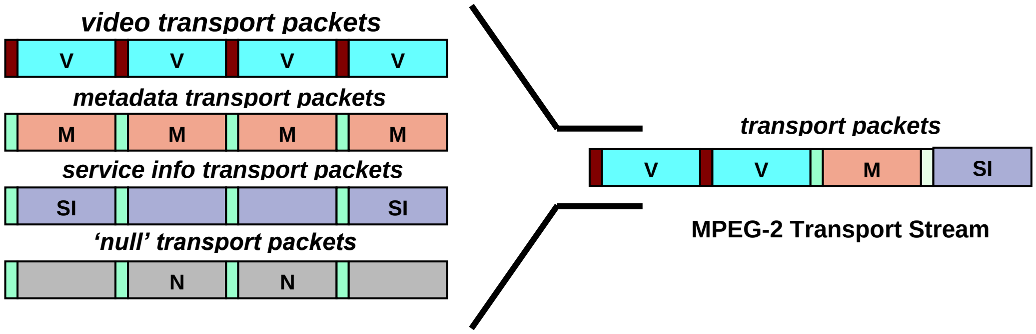


Figure 5: Step C: mapping PES packets to transport stream packets

Finally, in Step D (Figure 6) the transport stream multiplexer ingests the individual packetized media streams into transport-stream-format packets, and interleaves them into one composite stream of data. The data rate of the final TS stream will reflect the aggregate sum of respective individual media components. The ‘Null’ packet serves to fill in gaps in the stream to help maintain a desired transport stream data rate.



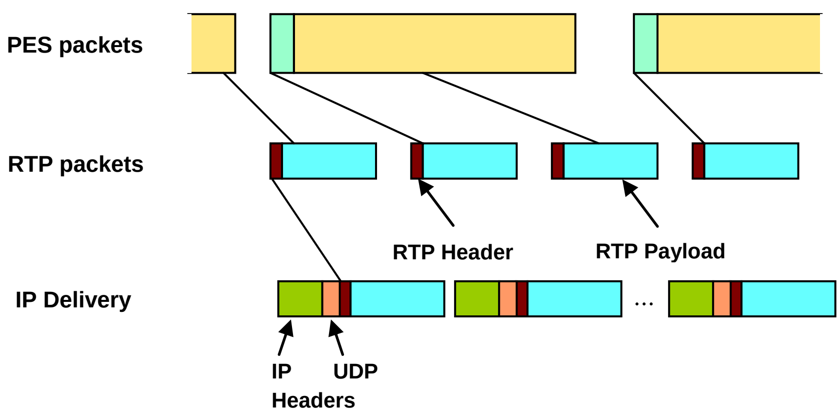
Transport Stream composed of a mix of various elementary streams

Figure 6: Step D: multiplexing media transport packets into one stream

Although the process above illustrates how video is mapped into the transport stream, the metadata is similarly packetized and receives a PES header. Once packets of metadata are constructed they are then mapped into transport stream packets with a corresponding PTS if the streams are intended to be displayed together with tight timing constraints.

8 Real Time Transport Protocol

The packaging of media for RTP follows the same first two steps of the process above for transport stream. Figure 7 shows a different step where the packetized data (PES packets) receive a RTP header on each PES packet, which then is encapsulated into an IP datagram with IP and UDP headers. A new draft for carriage of metadata over RTP is now under review by the Internet Engineering Task Force (IETF), a governing body responsible for publications on RTP usage. Reference is made to *RTP Payload for SMPTE 336M Encoded Data*, draft-ietf-avt-rtp-klv-00.txt [20]. For more information on RTP and a prototype implementation see **MISB RP 0804 - Real-Time Protocol for Full Motion Video [21]**.



IP, UDP, RTP headers precede data composed of PES packets

Figure 7: Packaging one media component for RTP delivery

9 File/Stream Protocol

The motion imagery essence and the metadata together represent the collected “asset”. To be of value this asset must be kept intact within either a unified container, or with appropriate timing, to facilitate downstream decisions. There are several packaging methods available within the MISB suite of technologies: MPEG-2 TS (Transport Stream), RTSP/RTP, and AAF/MXF. These different packages are useful in different applications. MPEG-2 TS is the most used container in our community. It serves as both a vehicle for point-to-point transport and also for storage.

10 Advice on Getting Started

The MISB web site (http://www.gwg.nga.mil/misb/index.html) publishes all the documents needed to construct a MISP-compliant file or stream. The MISP is the authoritative MISB reference. It contains references to all the Standards (STD), Recommended Practices (RP), Engineering Guidelines (EG), and Technical Reference Material (TRM) supported by the MISB, and also provides up-to-date industry references foundational to the MISB documents. The series of MISM (Motion Imagery System Matrix) tables for electro-optical imagery and IRSM (Infrared System Matrix) tables for Infrared imagery found in the MISP provide guidance on motion imagery formats, compression rates, transport channels and protocols.

The relevant MISB documentation that supports the example implementation above is recapped in Table 3 and denoted in Figure 8 (red markings). By reviewing these references, in concert with the MISP, no critical elements necessary for MISP compliance should be missed.

Electro-Optical Motion Imagery Format	MISB STD 9710 **MISB RP 9720b**	High Definition Television Systems (HDTV) **MISM, High Definition Motion Imagery**
Transport Protocol	MISB STD 9703	Digital Motion Imagery, Uncompressed Baseband Signal Transport and Processing
Compression	**MISB STD 9723** MISB EG 0802	**Compressed High Definition Advanced Television (ATV) and associated Motion Imagery Systems** H.264/AVC Coding and Multiplexing
Metadata	MISB STD 0601 **MISB STD 0102** **MISB STD 0902** MISB STD 0807 MISB STD 0107 MISB STD 9715 MISB STD 9713 **MISB RP 9717** MISB EG 0607	UAS Datalink Local Metadata Set **Security Metadata Universal and Local Sets for Digital Motion Imagery** **Motion Imagery Sensor Minimum Metadata Set** MISB KLV Metadata Dictionary Bit and Byte Order for Metadata in Motion Imagery Files and Streams Time Reference Synchronization Data Encoding Using Key-Length-Value **Packing KLV Packets into MPEG-2 Systems Streams** MISB Metadata Registry and Processes
Timing/Multiplexing	MISB STD 9708 **MISB STD 0604** MISB RP 0603 **MISB RP 0605**	Imbedded Time Reference for Motion Imagery Systems **Time Stamping and Transport of Compressed Motion Imagery and Metadata** Common Time Reference for Digital Motion Imagery Using Coordinated Universal Time (UTC) **Inserting Time Code and Metadata in High Definition Uncompressed Video**
Media Container	MISB STD 9601 **MISB STD 9701** MISB RP 0101	Standard Definition Digital Motion Imagery, Compression Systems **MPEG-2 Transport Stream** Use of MPEG-2 System Streams in Digital Motion Imagery Systems

Table 3: Referenced MISB Documents for System in Figure 2

Various MISB references for compression, metadata, timing and transport protocol or storage can be quickly viewed with charts like that in Figure 8, found in **MISB TRM 1007 - Surfing the MISP [25]**. While there are many documents, not all are necessary for a given application. Of most importance is diligence in following the particular metadata standard or practice—particularly security as outlined in STD 0102 and the required metadata in STD 0902. STD 0601 is instructive in understanding how KLV packets are encoded in KLV, and other topics like platform data flow and error detection. STD 0604 addresses time stamping and time code for the motion imagery, time stamping of the metadata, synchronization of the imagery with the metadata, and how best to carry the metadata along with the motion imagery.

A motion imagery asset is much more valuable when it is accompanied with metadata; collecting pure MI alone is of little use downstream for exploitation. MISP compliant streams must contain metadata. Mission metadata is critical for assisting discovery and retrieval of an asset, and for relating one asset or intelligence data to another. The MISB provides encompassing metadata sets to support mission metadata. How to add enriched metadata, such as annotation, chat, audio, etc. is also provided through various MISB RP's and EG's. As more motion imagery is manipulated throughout the network with processes—such as transcoding for bandwidth-challenged edge delivery—the MISB will continue to look to industry for best practices and adopt or leverage those found applicable.

It is advisable to revisit the MISB web site periodically for updates and additions that may affect implementation decisions. The MISB encourages community participation and feedback in shaping the recommendations to best fit community needs. The goal is interoperability with standards as foundational and clarity an objective in our work.

The MISB supplies test files that represent various compression and metadata combinations for both standard and high definition video. These First Level Integration (FLI) files are generally less than one minute in duration. The metadata is typically random with respect to the actual video content. Although random, the metadata exercises all the keys and tags of a particular MISB metadata set with data values that are permitted. FLI test files are available on the protected MISB site for community use in testing products and systems.

Lastly, MISP testability for compliance is to a particular version of the MISP with certifications issued by the JITC (Joint Interoperability Test Command). Questions or misinterpretations of MISB documentation should be brought to the attention of the MISB. Periodic meetings (three per year) are held to discuss such issues and provide appropriate changes where needed.

Contact information for the MISB and the JITC:

MISB: http://www.gwg.nga.mil/misb/contact.html

JITC: Joint Interoperability Test Command/MIS-LAB
Fort Huachuca, AZ 85670-2798
Phone: (520) 538-1888
E-mail: MIS-LAB@disa.mil Website URL: http://jitc.fhu.disa.mil/mis/

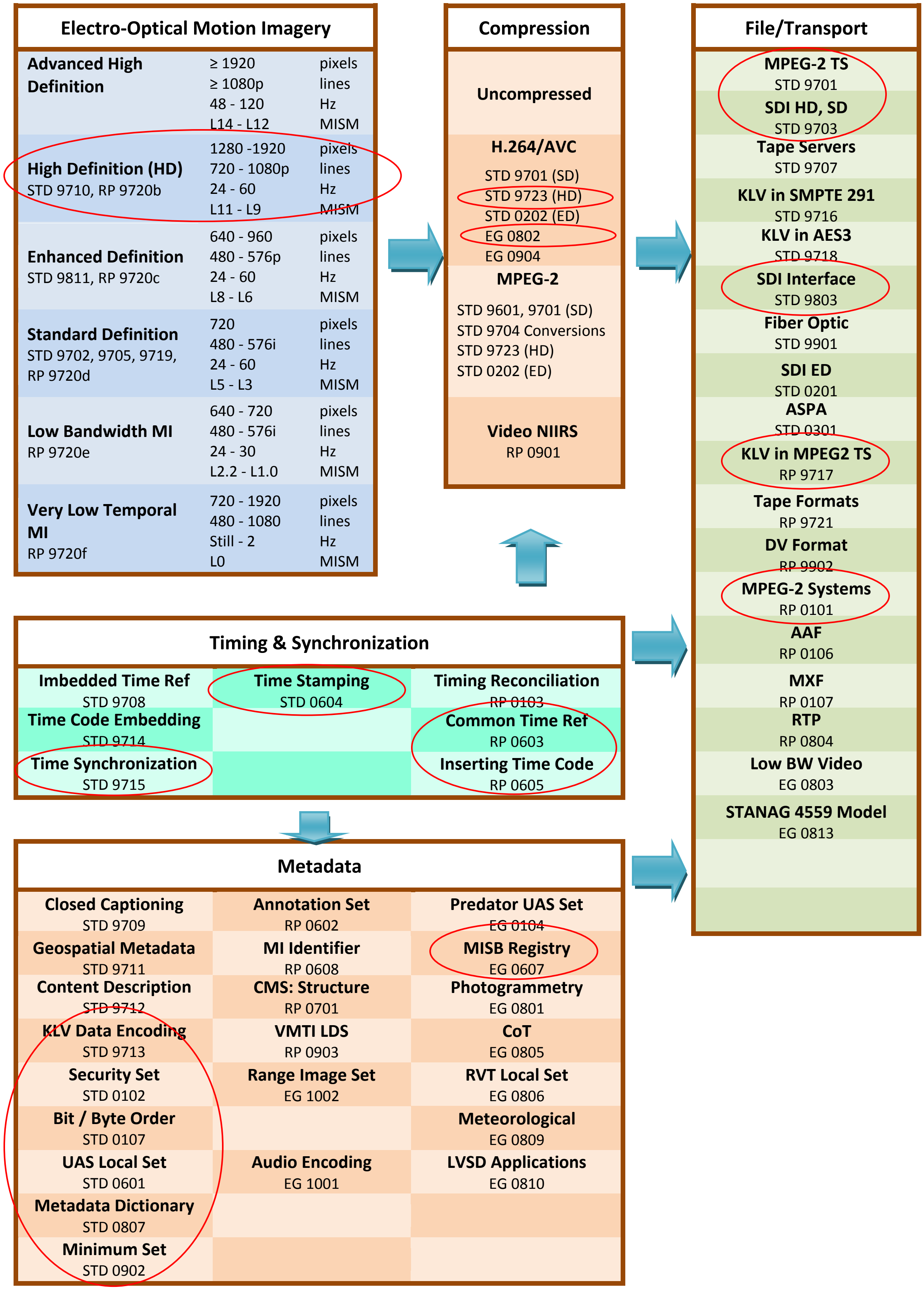


Figure 8: Referenced MISB Documents (in red) for System in Figure 2

Appendix A – An Analog NTSC Sensor (informative only)

While use of analog in any new deployments is strongly discouraged this section is intended to provide continuity in understanding of the processes from legacy system to digital system implementations.

Figure A1 illustrates a simplified sensor collection system where analog video is captured along with mission metadata. In this example, it is assumed that the encoding, metadata processing, and container packaging is done at the ground station.

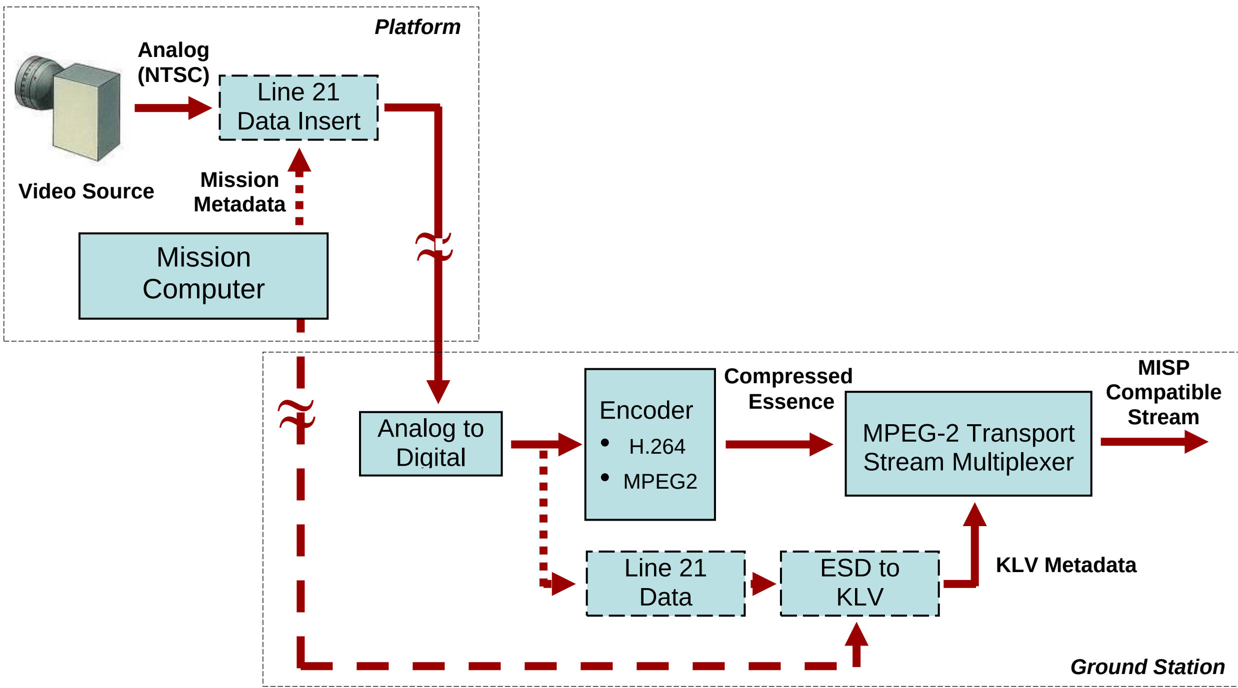


Figure A1: Analog Video Sensor System

Mission metadata may be inserted into the analog video on Line 21 (designated for television closed captioning), or transmitted directly to the ground station through an independent path. Once at the ground station, the analog video is converted to digital and then encoded using H.264/AVC or MPEG-2 compression. Metadata (in ESD format) inserted on Line 21 at the platform will be extracted and converted to KLV. Metadata that arrives via an independent path from the platform will likewise be converted to KLV. In both cases the KLV metadata will be multiplexed with the compressed video essence into a MPEG-2 Transport Stream container for dissemination or storage.

Let's examine some of the elements in this process further. The NTSC analog television standard supports 525 lines for one complete video frame. This is delivered to the display as two successive "fields," where each field is 262.5 lines, and reconstructed as complete frames in a process called interlace. Think of a numbered list of lines that make up an image where in one field the odd lines are "painted" followed by a second field that paints the even lines. Because the display phosphors have persistence and retain a portion of the previous field the illusion to the viewer is a flicker-free rendition of the picture.

The NTSC standard allows for roughly 482 lines to be used for picture content with the remaining lines—called the Vertical Blanking Interval (VBI)—hidden behind the bezel of the display. Line 21 is one such unused picture line. With the introduction of closed captioning Line 21 became one of the designated places to carry this data. This is also where metadata is often placed in analog capture systems. Mission metadata is sometimes modulated onto Line 21 as a means to carry the data with the video.

At the ground station the analog video is converted to digital through an analog-to-digital converter. This produces a number of samples called pixels (picture elements) horizontally. The vertical direction is already "digital" in that each scan line is discrete and independent of adjacent ones. Each pixel is typically quantized to eight bits of gray-scale resolution and the color components may be quantized to eight but typically fewer bits, such as four.

The digitized video signal is ready for compression through the encoder. In parallel, Line 21 of the analog video is stripped of any metadata that may have been inserted at the platform and converted to KLV format. Any metadata that comes directly from the platform independent of the video is also converted, if necessary. All KLV metadata is finally input to the transport stream multiplexer along with the compressed video essence.

The function of the transport stream multiplexer is to insert the two (or more) different data sources into one common package. One can think of the transport stream as containing a series of parallel tracks where each track can carry its own data type—analogous to a multi-lane highway. The "multiplex" serves to merge the various data lanes into one—think of merging traffic at an on-ramp. Certain track identifiers and management directories are placed into the transport stream to guide the extraction of the various data types at the decoder.

Similarly to the digital FMV example in the text optional metadata may be added to the composite multiplexed MI file/stream anywhere along the PED workflow. Such metadata may include chat, audio, annotations and other data types. See the MISB website http://www.gwg.nga.mil/misb/index.html for documents that guide such metadata additions.

Appendix B – TS, PES Headers (informative only)

Tables B1 and B2 list various flags and data fields that comprise the transport stream (TS) header and the packetized elementary stream (PES) header (ISO/IEC 13818-1 [26] Tables 2-2, 2-21) respectively. Required values for the carriage of metadata as indicated in STD 0604 – *Time Stamping and Transport of Compressed Motion Imagery and Metadata* are shown in the metadata-only tables (Tables B3, B4), as are values that although optional are recommended for consistency in usage.

Table B1: Transport Stream Header ***(4 bytes minimal)***			
	No. of bits	***shall***	***recommended***
sync_byte	8	0x47	
transport_error_indicator	1		usage based
payload_unit_start_indicator	1		usage based
transport_priority	1		'0'
PID	13		usage based
transport_scrambling_control	2	'00' (not scrambled)	
adaptation_field_control	2		usage based
additional flags, data possible	K bytes		usage based
continuity_counter	4		usage based

Table B2: Packetized Elementary Stream Header			
	No. of bits	***shall***	***recommended***
packet_start_code_prefix	24	0x000001	
stream_id	8		usage based
PES_packet_length	16		usage based
'10'	2	'10'	
PES_scrambling_control	2	'00' (not scrambled)	
PES_priority	1		'0' (low)
data_alignment_indicator	1		usage based
copyright	1		'0' (not specified)
original_or_copy	1		'1' (original)
PTS_DTS_flags	2		usage based
ESCR_flag	1		usage based
ES_flag_rate	1		usage based
DSM_trick_mode_flag	1	'0'	No trick mode field
additional_copy_info_flag	1		usage based
PES_CRC_flag	1		usage based
PES_extension_flag	1		usage based
PES_header_data_length	8		usage based
If (PTS_DTS_flags == '10')			
'0010'	4	'0010'	fixed
PTS[32..30]	3		usage based
marker_bit	1	'1'	
PTS[29..15]	15		usage based
marker_bit	1	'1'	

Table B2: Packetized Elementary Stream Header			
PTS[14..0]	15		usage based
marker_bit	1	'1'	fixed
If (PTS_DTS_flags == '11')			
'0010'	4	'0010'	fixed
PTS[32..30]	3		usage based
marker_bit	1	'1'	fixed
PTS[29..15]	15		usage based
marker_bit	1	'1'	fixed
PTS[14..0]	15		usage based
marker_bit	1	'1'	fixed
'0001'	4	'0001'	fixed
DTS[32..30]	3		usage based
marker_bit	1	'1'	fixed
DTS[29..15]	15		usage based
marker_bit	1	'1'	fixed
DTS[14..0]	15		usage based
marker_bit	1	'1'	fixed
If ESCR_flag == '1'			
Reserved	2	'11'	reserved
ESCR_base[32..30]	3		usage based
marker_bit	1	'1'	fixed
ESCR_base[29..15]	15		usage based
marker_bit	1	'1'	fixed
ESCR_base[14..0]	15		usage based
marker_bit	1	'1'	fixed
ESCR_extension	9		usage based
marker_bit	1	'1'	fixed
If (ES_rate_flag == '1')			
marker_bit	1	'1'	fixed
ES_rate	22		usage based
marker_bit	1	'1'	fixed
If (additional_copy_info_flag == '1')			
marker_bit	1	'1'	fixed
additional_copy_info	7		usage based
If (PES_CRC_flag == '1')			
previous_PES_packet_CRC	16		usage based
If (PES_extension_flag == '1')			
PES_private_data_flag	1		usage based
pack_header_field_flag	1		usage based
program_packet_sequence_counter_flag	1		usage based
P-STD_buffer_flag	1		usage based
Reserved	3	'111'	reserved
PES_extension_flag_2	1		usage based
If (PES_private_data_flag == '1')			

Table B2: Packetized Elementary Stream Header			
PES_private_data	128		usage based
If (pack_header_field_flag == ‘1’)			
pack_field_length	8		usage based
If (program_packet_sequence_counter_flag == ‘1’)			
marker_bit	1	‘1’	fixed
program_packet_sequence_counter	7		usage based
marker_bit	1	‘1’	fixed
MPEG1_MPEG2_identifier	1		usage based
original_stuff_length	6		usage based
If (P-STD_buffer_flag == ‘1’)			
‘01’	2	‘01’	fixed
P-STD_buffer_scale	1		usage based
P-STD_buffer_size	13		usage based
If (PES_extension_flag_2 == ‘1’)			
marker_bit	1	‘1’	fixed
PES_extension_field_length	7		usage based
stream_id_extension_flag	1		usage based
If (stream_id_extension_flag == ‘0’)			
stream_id_extension	7		usage based
data	N x 8		usage based
stuffing_byte	N1 x 8		usage based
PES_packet_data_byte	N2 x 8		usage based

Table B3: PES Header for Asynchronous Metadata Carriage			
	No. of bits	***shall***	***recommended***
packet_start_code_prefix	24	0x000001	
stream_id	8	0xBD (private stream 1)	
PES_packet_length	16		usage based
‘10’	2	‘10’	
PES_scrambling_control	2	‘00’ (not scrambled)	
PES_priority	1		‘0’ (low)
data_alignment_indicator	1		usage based
copyright	1		‘0’ (not specified)
original_or_copy	1		‘1’ (original)
PTS_DTS_flags	2	‘00’ (no PTS/DTS)	
ESCR_flag	1	‘0’ (not present)	
ES_flag_rate	1	‘0’ (not present)	
DSM_trick_mode_flag	1	‘0’ (not present)	
additional_copy_info_flag	1	‘0’ (not present)	
PES_CRC_flag	1	‘0’ (not present)	
PES_extension_flag	1	‘0’ (not present)	
PES_header_data_length	8	N1 (bytes)	
stuffing_byte	N1 x 8		N1 bytes
PES_packet_data_byte	N2 x 8		N2 bytes

Table B4: PES Header for Synchronous Metadata Carriage			
	No. of bits	shall	recommended
packet_start_code_prefix	24	0x000001	
stream_id	8	0xFC (metadata stream)	
PES_packet_length	16		usage based
‘10’	2	‘10’	fixed
PES_scrambling_control	2	‘00’ (not scrambled)	
PES_priority	1		‘0’ (low)
data_alignment_indicator	1		usage based
copyright	1		‘0’ (not specified)
original_or_copy	1		‘1’ (original)
PTS_DTS_flags	2	‘10’ (PTS only)	
ESCR_flag	1	‘0’ (not present)	
ES_flag_rate	1	‘0’ (not present)	
DSM_trick_mode_flag	1	‘0’ (not present)	
additional_copy_info_flag	1	‘0’ (not present)	
PES_CRC_flag	1	‘0’ (not present)	
PES_extension_flag	1	‘0’ (not present)	
PES_header_data_length	8	5+N1 (bytes)	usage based
If (PTS_DTS_flags == ‘10’)		‘10’ (PTS only)	
‘0010’	4	‘0010’	fixed
PTS[32..30]	3		usage based
marker_bit	1	1	fixed
PTS[29..15]	15		usage based
marker_bit	1	1	fixed
PTS[14..0]	15		usage based
marker_bit	1	1	fixed
stuffing_byte	N1 x 8		N1 bytes
PES_packet_data_byte	N2 x 8		N2 bytes